

日本語表記選択支援システムの研究

大谷 紀子研究室

0432058 加藤浩郷

1. 研究の背景と目的

日本語には様々な異なった表記を持つ言葉が多く存在する。複数ある表記の中から適切な一つの表記を選択することは難しい。インターネット上では、様々な表記で情報が公開されている。人々が独自に作成したもので、一般的な使用傾向が反映されているので、適切な表記を選択するための材料として利用できると考えられる。本研究では日本語の複数の表記を持つ言葉について、ユーザが適切な表記を選択する際の支援を目的とする。複数の表記が存在する言葉に関して、インターネット上の情報から得られる一般的な使用傾向を表示することで、適切な表記選択を支援する機能を提案する。本機能を搭載したシステムを構築し、評価実験によりシステムの有用性を示す。

2. システム概要

日本語表記の選択支援の需要を調査するために、武蔵工業大学の3年生と4年生の計35人に普段文章を書く際の表記選択に関するアンケートをとった。アンケートの結果を図1、図2に、内容を以下に示す。

質問1 文章を書く際の表記について迷うことはあるか。

質問2 どのような表記に迷うか。特にどんな場合か。

- 1 漢字か平仮名か
- 2 カタカナか平仮名か
- 3 漢字、平仮名、カタカナの組み合わせ
- 4 送り仮名
- 5 濁点の有無
- 6 伸ばし棒の有無
- 7 発音とのゆれ
- 8 同義の漢字

図1から29人が迷うことがあるという結果が得られ、図2では、大小はあるが全ての場合について迷う人がいるという結果が得られた。日本語表記のゆれに関する選択支援の需要はとても高いといえる。

本研究ではユーザの表記検索を支援するシステムを構築する。表記を選択する際に、各表記の使用傾向を表示することでユーザの選択を支援する。インターネット上での各表記の使用頻度を人々の使用傾向と捉え、インターネット検索から得られた各表記の検索結果数を表示する。

システムの流れを以下に示す。

- (1) ユーザは調べる表記を入力する。
- (2) Yahoo!検索APIから返された検索結果数が表示される。

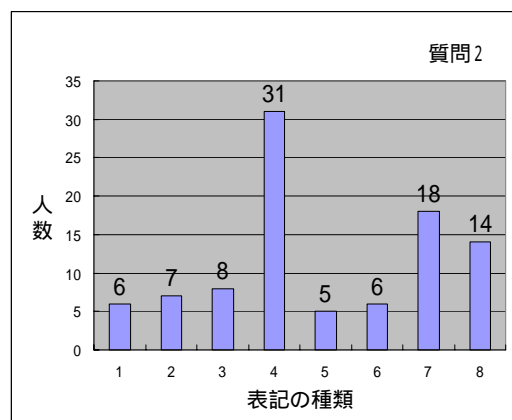
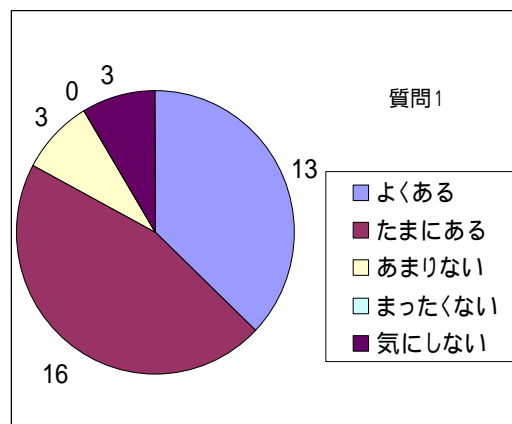


図2：難しい表記選択

(3) 別の表記と比較する場合は(1)へ戻り、結果を表示する場合は(4)へ進む。

(4) 今まで検索した結果が並べて表示される。

結果画面では検索結果数を数字とグラフの両方で表示するので、ユーザが各表記の検索結果数の大小を簡単に比較することができる。結果画面を図3の例に示す。



図3：システム結果画面

3. 評価実験

武蔵工業大学の3年生、4年生35人を被験者として、評価実験を行った。短文を用意し、文中の指定した言葉の表記を選択肢から選ばせる問題を2問出す。設問1は本システムを使わずに、被験者の正しいと思う表記を選ばせ、設問2では本システムの表示する結果を考慮したうえで表記を選択させる。回答後、システムの有用性に関するアンケートを取った。アンケート結果を図4に示す。

4. 考察

有用性についてのアンケートの結果、図1で表記を気にしないと答えた3人が「思わない」と回答した以外は、「使ってみたいと思う」、または「時々なら」という回答が得られ、本システムの有用性を十分示すことができたと思われる。しかし、10人の被験者が「時々なら」と回答している点、および30%の被験者から挙げられた「検索結果数の信頼性に不安がある」という意見に関しては再考の余地がある。評価実験の設問2では、本システムの示す結果から最も検索結果数の多い表記を回答すると、全員同じ表記を選択するはずだが、結果にはばらつきがあり、被験者が結果を信頼していない可能性がある。また、検索結果数が使用傾向を表しているとしても、多く使用されているから正しいとはいえないという意見もあった。本システムでは日本語表記の正誤を考慮していないため、使用傾向と表記の正しさの相関については言及できない。しかし、正誤を示すシステムでないことはあらかじめ伝えたとうえで、役に立つという回答が多かったことから、検索結果数が多ければ正しいとはいえなくても、表記選択の材料にはなると考える。実際に、表記の正誤を求められるような文書でなければ使ってみたいという意見もあった。結果の表示の仕方については、グラフ表示が見やすいという意見が多くあり、選択の支援に役立つと考える。今後の課題としては、表記の一括入力、前に入力した表記の表示など使いやすさにおける改善点が挙げられる。また日本語の言い回しなどの比較にも使えるという意見もいくつか挙げられた。本研究では日本語表記の選択支援のみを対象としているが、今後は言い回しや外国語など、日本語表記以外の選択の支援にもインターネットでの検索結果数が役立つ可能性があると考えられる。

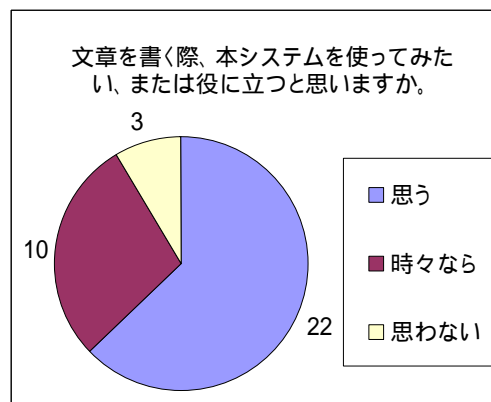


図4：システムの有用性